第 16 章 主成分分析和因子分析

C.Radhakrishna Rao

1.引言

主成分分析和因子分析（principal component and factor analysis,PCA 和 FA）是用于研究个体测量指标的协方差（或相关）结构的探索性多元技术。分析的目标可以不同:找出若干可以解释可观测指标之间的变差或者联系的潜变量而简化高维数据,对相似指标进行分组及检测多重共线性，将高维数据在低维空间中图示以直观考察数据的散布情况及检测异常值。PCA 是由 Pearson（1901）和 Hotelling（1933）发展起来的；Rao（1964）给出了基本原理及一些扩展和运用。FA 首先由 Spearman（1904）提出，接着 Lawley（1940）在多元正态性的假定下发展。Rao（1955）给出了没有任何分布假设的 FA 的原理，命名为典型因子分析（CFA）。现在已经有许多优秀的、大部的专著致力于社会科学和自然科学研究中 PCA 和 FA 的计算及使用问题。参考文献包括 Bartholomew（1987），Basilevsky（1994），Cattel（1978），Jackson（1991），和 Jolliffe（1986），这里仅提到少数作者。

当测量指标是定性指标时，和 PCA 有关的一种方法称为对应分析（correspondence analysis,CA），是 Benzecri（1973）基于 Fisher（1936）提出的定性尺度范畴（scaling qualitative categories）方法而发展起来的。Greenacre（1984）的专著阐述了 CA 的理论及其在列联表分析中的应用。Rao（1995）的论文包含了 CA 的一种替代方法，和 CA 的用途一致，但看起来优于早期方法。

本文将提供某些最新理论成果和实际应用以全面考察 PCA 和 FA。

2.主成分

2.1.一般问题

主成分问题可以用如下非常一般的步骤进行阐述。令 x 是一个 p 维向量， y 是一个 q 维向量，其中 x 和 y 的一些分量可能是相同的。我们要用 $z = Ay$ 代替 y，其中 A 是一个 $r \times q$ 矩阵且 $r < q$，使得用 z 代替 y 预测 x 的损失尽可能小。如果

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.1}$$

是 x 和 y 的协方差矩阵，那么用 $z = Ay$ 预测 x 的误差的协方差矩阵就是

$$W = \Sigma_{11} - \Sigma_{12} A' \left(A \Sigma_{22} A' \right)^{-1} A \Sigma_{21} \tag{2.2}$$

我们选择 A，使得 $\|W\|$ 对一个适当选择的标准而言是小的。如果选择 $\|W\| = \mathrm{tr} W$，那么最优选择是

$$A_* = \arg\max_A \mathrm{tr} \Sigma_{12} A' \left(A \Sigma_{22} A' \right)^{-1} A \Sigma_{21}$$

最大值在

$$A_*' = \left(C_1 : \cdots : C_r\right) \tag{2.3}$$

时达到，其中$C_1,\cdots,C_r$是$\Sigma_{21}\Sigma_{12}$关于Σ_{22}的前r个特征值$\lambda_1^2 \ge \lambda_2^2 \ge \cdots \ge \lambda_r^2$所对应的$r$个特征向量，也就是说，特征向量和特征值是从下面的行列式方程中产生的

$$\left|\Sigma_{21}\Sigma_{12} - \lambda^2\Sigma_{22}\right| = 0 \tag{2.4}$$

用$z_* = A_* y$预测x的相对信息损失是

$$\begin{aligned} &\mathrm{tr}\left(\Sigma_{11} - \left(A_*\Sigma_{22}A_*'\right)^{-1} A_*\Sigma_{21}\Sigma_{12}A_*'\right) / \,\mathrm{tr}\Sigma_{11} \\ &\quad = 1 - \frac{\lambda_1^2 + \cdots + \lambda_r^2}{\mathrm{tr}\Sigma_{11}} \end{aligned} \tag{2.5}$$

我们可以选择一些特殊的x和y，并导出具有（2.3）形式的最优变换A。

2.2. x=y 的情形

考虑特定情形$x = y$，得到一般主成分$C_1'x,\cdots,C_r'x$，其中$C_1,\cdots,C_r$是行列式方程$\left|\Sigma_{11} - \lambda I\right| = 0$的前$r$个特征值$\lambda_1^2 \ge \cdots \ge \lambda_r^2$所对应的前$r$个特征向量。在这种情况下，信息损失（2.5）等于

$$1 - \frac{\lambda_1^2 + \cdots + \lambda_r^2}{\lambda_1^2 + \cdots + \lambda_p^2} = \frac{\lambda_{r+1}^2 + \cdots + \lambda_p^2}{\lambda_1^2 + \cdots + \lambda_p^2} \tag{2.6}$$

通常用百分数表示，r的选择取决于（2.6）的大小。

实际上，需要从p维随机向量x的n个独立观测值样本中估计λ_i^2和C_i，这个样本用$p \times n$矩阵表示

$$X = \left(x_1 : \cdots : x_n\right) \tag{2.7}$$

Σ_{11}的估计量是

$$S = (n-1)^{-1} X\left(I - \frac{1}{n}ee'\right)X'$$

其中e是n维 1 向量。λ_i的估计ℓ_i和C_i的估计量c_i可以从谱分解中得出

$$S = \ell_1^2 c_1 c_1' + \cdots + \ell_p^2 c_p c_p' \tag{2.8}$$

于是第i个个体观测值的主成分就是

$$q_i = \left(c_1' x_i, \cdots, c_p' x_i\right)' \tag{2.9}$$

其结果是，我们表示

$s_{ii} = S$的第i个对角线元素，

$$c_j = \left(c_{j1}, \cdots, c_{jp}\right)', j = 1, \cdots, p, \tag{2.10.1}$$

$$\hat{c}_{ji} = \ell_j c_{ji}, i = 1, \cdots, p, \tag{2.10.2}$$

$$q_i = \left(q_{i1}, \cdots, q_{ip}\right)', i = 1, \cdots, n, \tag{2.11.1}$$

$$\hat{q}_{ij} = \ell_j^{-1} q_{ij}, i = 1, \cdots, n. \tag{2.11.2}$$

应该注意，向量 c_i 和 q_i（除坐标平移之外）可以从奇异值分解（singular value decomposition, SVD）

$$X\left(I - \frac{1}{n}ee'\right) = \ell_1 c_1 d_1' + \cdots + \ell_p c_p d_p' \tag{2.12}$$

一步得出。其中具有关系式 $\left(l_1 d_1 : \cdots : l_p d_p\right)' = \left(q_1 : \cdots : q_n\right)$。

2.3.主成分解释

为了用原始测量指标的影响解释主成分，需要表 1 中列出的计算。

表 1

原始变量	与主成分的相关系数			x_i 对 $z_1, \cdots, z_r$ 的复相关系数
	z_1	$\cdots$	z_p	
x_1	$\hat{c}_{11} / \sqrt{s_{11}}$	$\cdots$	$\hat{c}_{p1} / \sqrt{s_{11}}$	$s_{11}^{-1} \sum_{j=1}^{r} \hat{c}_{j1}^2 = R_1^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_p	$\hat{c}_{1p} / \sqrt{s_{pp}}$	$\cdots$	$\hat{c}_{pp} / \sqrt{s_{pp}}$	$s_{pp}^{-1} \sum_{j=1}^{r} \hat{c}_{jp}^2 = R_p^2$

表 1 中相关系数的大小表明每一个变量由每一个主成分代表的效果，以及全部变量用前 r 个主成分代表的效果（通过 R_i^2 的值判断）。对 $r = 1, 2, \cdots$ 计算 R_i^2 的值使得我们能够对 r 作出决定，即决定选取主成分的个数。如果对某个 r ,除一个 i 值以外 R_i^2 的值都很高，比如说是 j ，那么可以决定将 x_j 加入 $z_1, ..., z_r$ 中，或者增加其他能很好代表 x_j 的主成分。

2.4.数据的图形显示

依照原始测量指标表示所有个体，需要一个 p 维空间。但是为了直观地考察，我们需要这些个体在二维或者三维空间里的标示图，它应该尽可能反映个体在 p 维空间里的构形 (configuration)（个体之间的距离）。为此，使用（2.11.1）中的主成分或者（2.11.2）中标准化形式的主成分（SPC）。表 2 列出了完整的各种维数的新坐标，可以选择前面少量维数。

表 2

个体	一维		二维		$\cdots$	p 维	
	PC	SPC	PC	SPC	$\cdots$	PC	SPC
1	q_{11}	$\hat{q}_{11}$	q_{12}	$\hat{q}_{12}$	$\cdots$	q_{1p}	$\hat{q}_{1p}$
2	q_{21}	$\hat{q}_{21}$	q_{22}	$\hat{q}_{22}$	$\cdots$	q_{2p}	$\hat{q}_{2p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	q_{n1}	$\hat{q}_{n1}$	q_{n2}	$\hat{q}_{n2}$	$\cdots$	q_{np}	$\hat{q}_{np}$
方差	ℓ_1^2	1	ℓ_2^2	1	$\cdots$	ℓ_p^2	1

如果对所有个体,第 i 个个体只取它的前 r （$< p$）个坐标 $q_{i1},\cdots,q_{ir}$，把它们投射到 r（$< p$）维空间中，那么第 i 个和第 j 个个体之间的欧氏（Euclidean）距离将是全 p 维空间中欧氏距离

$$d_{ij}=\left[\left(x_i-x_j\right)'\left(x_i-x_j\right)\right]^{1/2}$$

的近似值。

另一方面，如果用坐标 $\hat{q}_{i1},\cdots,\hat{q}_{ir}$ 将所有个体表示在前 r（$< p$）维空间中，那么在这样的空间中个体 i 和 j 之间的欧式距离将是 p 维空间中马氏（Mahalanobis）距离

$$d_{ij}=\left[\left(x_i-x_j\right)'S^{-1}\left(x_i-x_j\right)\right]^{1/2}$$

的近似值。

实际上，需要选择要在降维空间中保存的适当距离。通常，二维或者三维图就足够反映原始构形。如果需要大于三维的空间，就要使用其他可以使高维图形象化的显示方法了。例如，参见 Wegman，Carr 和 Luo（1993）的论文。

我们也可以将变量表示在低维空间中，以直观考察变量之间的联系。表 3 列出了用于这一研究目标的全部坐标。

表 3

变量	坐标			
1	$\hat{c}_{11}$	$\hat{c}_{21}$	$\cdots$	$\hat{c}_{p1}$
2	$\hat{c}_{12}$	$\hat{c}_{22}$	$\cdots$	$\hat{c}_{p2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
p	$\hat{c}_{1p}$	$\hat{c}_{2p}$	$\cdots$	$\hat{c}_{pp}$

我们把 r 维空间中代表第 i 个个体的点和原始变量联系起来的向量记为 v_i。那么 $v_i'v_i$ 就是 s_{ii} 的一个良好近似值，第 i 个变量的方差和向量 v_i、v_j 夹角的余弦将是第 i 与第 j 个变量之间相关系数的一个良好近似值。

2.5.*残差分析和异常值检测*

如果保留前 r 个主成分，可以计算用 $\hat{x}_i$ 近似 x_i（第 i 个个体的 p 维测量指标向量）的误差，

即

$$x - \hat{x}_i = \left(c_{r+1}c'_{r+1} + \cdots + c_p c'_p\right)x$$

而度量的总误差是

$$d_i^2 = \left(x_i - \hat{x}_i\right)'\left(x_i - \hat{x}_i\right) = q_{ir+1}^2 + \cdots + q_{ip}^2$$

如果某个 d_i^2 相对其他而言较大，就表明 x_i 可能是一个异常值。

要点 1. 主成分并非对原始变量的线性变换具有不变性。例如，如果将原始变量以不同的数按比例调整或者进行旋转（rotated）线性变换，那么得出的主成分将是不同的。这意味着必须进行初始决策（initial decision），将原始测量指标变换为一组新的测量指标，然后再求取主成分。通常的建议是将测量指标除以标准差，这相当于主成分是从相关矩阵求得的而不是从协方差矩阵求得。

要点 2. 当原始测量指标服从多元正态分布时，有许多方法可用于协方差矩阵的特征值与特征向量的检验[参见 Basilevsky（1994）的第四章]。实践中，如果要应用这些检验，就必须进行原始测量指标的正态性检验。必要的话，可以采用 Box-Cox 变换族对测量指标进行变换以引致正态性。有几个计算机程序具有这种选项。在这种情况下，计算的是变换后变量的主成分。

要点 3. 在某些问题中，如分析增长曲线，主成分是从没有进行均值校正的矩阵 $S = XX'$ 求出。这种方法可参考 Rao（1958，1987）。

要点 4. Jolicoeur 和 Mosimann（1960）提出，具有最大方差的第一主成分（如果所有系数都为正）可以解释为规模因子(size factors)，而系数有正有负的其他主成分可以解释为形状因子(shape factors)。这种解释的一个理由可能是这样的：考虑 x 中的第 i 个变量 x_i 和 x 的第 j 个主成分 $c'_j x$。x_i 对 $c'_j x$ 回归的系数是 c_{ji}，即第 j 个特征向量 c_j 中的第 i 个元素。现在 $c'_j x$ 增加 1 单位将使得 x_i 平均增加 c_{ji}。如果 c_j 中所有元素都为正，$c'_j x$ 增加 1 单位将增加每一个测量指标的值，在这种情况下 $c'_j x$ 被描述为规模因子。如果一些系数为正而其他系数为负，那么 $c'_j x$ 的增加将使得一些测量指标的值增加而其他测量指标的值减小，在这种情况下 $c'_j x$ 被描述为形状因子。

有趣的是注意到，如果所有原始测量指标都是非负的，那么从未校正平方和与积的矩阵求得的第一主成分的所有系数也将是非负的。

要点 5. 2.1 节中阐述的一般问题的另一个特例是，x 和 y 是完全不同的变量集。当有很多所谓的工具变量（用 y 表示），并且希望用 y 的某一线性函数来预测集合 x 中的每一个因变量时，就会出现这种情况。由于 y 中的多重共线性，这种过程可能更经济而且有时也更有效。

2.6. 与伴随变量(concomitant variable) z 无关的 x 的主成分

在某些问题中，人们感兴趣的是求 p 维向量 x 的主成分，它与伴随变量的 q 维向量 z 无关。令

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.13}$$

表示 $(x', z')'$ 的分块协方差矩阵。要求 k 个主成分 $L_1'x, \cdots, L_k'x$ 使得 $L_i'L_i = 1, L_i'L_j = 0$ 且 $\operatorname{cov}(L_i'x, z) = L_i'\Sigma_{12} = 0, i, j = 1, \cdots, k$，而

$$L_1'\Sigma L_1 + \cdots + L_k'\Sigma L_k \tag{2.14}$$

达到最大值。Rao（1964）证明了，$L_1, \cdots, L_k$ 的最优选择是下列矩阵

$$\left(I - \Sigma_{12}(\Sigma_{21}\Sigma_{12})^{-1}\Sigma_{21}\right)\Sigma_{11} \tag{2.15}$$

的前 k 个正交特征向量.作为一个应用，我们考虑 Stone（1947）研究的一个表示批量经济交易的 p 维时间序列向量。

经济交易	时期			
	1	2	$\cdots$	T
1	x_{11}	x_{12}	$\cdots$	x_{1T}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
p	x_{p1}	x_{p2}	$\cdots$	x_{pT}
时 间 伴 随 函数				
一次	1	2	$\cdots$	T
二次	1	2^2	$\cdots$	T^2

我们计算由主变量和伴随变量产生的 $(p+2)$ 阶的协方差矩阵，把 T 视为样本容量，

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{2.16}$$

其中 S_{11} 是 $p \times p$ 矩阵，S_{12} 是 $p \times 2$ 矩阵而 S_{22} 是 2×2 矩阵。

正交特征向量的必要数（necessary number）

$$\left(I - S_{12}(S_{21}S_{12})^{-1}S_{21}\right)S_{11} \tag{2.17}$$

使得 x 的主成分不受随时间变化的交易的线性趋势和二次趋势影响。通过适当选择伴随变量作为时间的幂就可能消除低阶或高阶趋势。

Stone（1947）研究了上面的问题，将具有内在经济意义的 x 的线性函数和那些表示时

间趋势及测量随机误差的函数隔离开。为此，他单独计算了变量 x 的协方差矩阵，并用 S_{11}（未涉及任何时间因素的分块矩阵)的特征向量求主成分。于是形成了识别主导主成分问题，它能解释大部分方差。这被解释成线性趋势，而其他主成分用经济术语进行解释。通常认为，用（2.17）中的矩阵求解主成分的方法更灵活，是消除任何阶趋势的更好方法，并且提供具有内在经济意义的线性函数。

3.基于主成分的模型

3.1.与因子分析模型相似的模型

假设个体 i 的 p 维测量指标向量 x_i 可以表示成

$$x_i = \alpha + Af_i + \varepsilon_i, i = 1, \cdots, n \tag{3.1}$$

其中 α 是 p 维向量，A 是为所有个体共有的 $p \times r$ 矩阵，f_i 是特定个体 i 的 r 维向量，ε_i 是随机向量，满足 $\mathrm{E}(\varepsilon_i) = 0$ 且 $V(\varepsilon_i) = \sigma^2 I, i = 1, \cdots, n$。模型（3.1）与 FA 模型相似，除了在 FA 模型中 ε_i 的协方差矩阵是元素可能不同的对角矩阵以外（参见本文的第 4 节）。我们考虑的问题是从模型（3.1）中估计一组 $A, f_i, \cdots, f_n$ 和 σ^2。注意结果不是唯一的，除非施加某种约束，如 A 的列是标准正交向量。可以将联合的模型（3.1）写成

$$X = \alpha e' + AF + E \tag{3.2}$$

其中 $X = (x_1 : \cdots : x_n)$ 是 $p \times n$ 矩阵，e 是 n 维 1 向量，而 F 是 $r \times n$ 矩阵。可以选择一个适当的标准，最小化

$$\| X - \alpha e' - AF \| \tag{3.3}$$

以估计 α，A 和 F。选择 Frobenius 标准将得到扩展的最小二乘法，其中表达式

$$\sum_{i=1}^{n} (x_i - \alpha - Af_i)' (x_i - \alpha - Af_i) \tag{3.4}$$

关于 α，A 和 $f_i, \cdots, f_n$ 求最小化。一个可能解（参见 Rao（1995））是

$$\hat{\alpha} = \bar{x}, \hat{A} = (c_1 : \cdots : c_r), \hat{f}_i = \hat{A}'(x_i - \bar{x}) \tag{3.5}$$

其中 $c_1, \cdots, c_r$ 是 $S = X(I - \frac{1}{n}ee')X'$ 的前 r 个特征向量。因而 $\hat{f}_i$ 是个体 i 的 r 个主成分向量。因此，可以得到和 2.2－2.5 节讨论的相同的解。σ^2 的估计量是

$$\hat{\sigma}^2 = \frac{n-1}{(n-r-1)(p-r)} \left(\ell_{r+1}^2 + \cdots + \ell_p^2 \right) \tag{3.6}$$

其中$\ell_{r+1}^2,\cdots,\ell_p^2$是$S$的最后$(p-r)$个特征值。

在某些问题中，可能适合于把模型（3.1）中的f_i看成一个随机变量，协方差矩阵是单位矩阵I。在这种情况下，

$$\mathrm{E}(S)=AA'+\sigma^2 I\text{，} \tag{3.7}$$

A的估计量是

$$\hat{A}=\left(\ell_1 c_1:\cdots:\ell_r c_r\right), \tag{3.8}$$

而σ^2的估计量是

$$\hat{\sigma}^2=\frac{n-1}{(n-r-1)(p-r)}\left(\ell_{r+1}^2+\cdots+\ell_p^2\right) \tag{3.9}$$

除了尺度因子外，这和（3.6）式相同。如果要估计（预测）f_i，可以使用f_i对x_i的回归，形式如下

$$\hat{f}_i=\hat{A}'\left(\hat{A}\hat{A}'+\hat{\sigma}^2 I\right)^{-1}\left(x_i-\bar{x}\right) \tag{3.10}$$

这不同于（3.5）的表达式。当要从几个具有相同设计矩阵(design matrix)的线性模型中同时估计参数时，就产生了类似情况。这种问题的讨论可以参见 Rao（1975）。

3.2.基于主成分模型的回归问题

对于$(p+1)$个维随机变量的向量(y,x)，我们有n个独立观测值，其中x是一个p维向量而y是一个纯量，

$$(y_1,x_1),\ldots,(y_n,x_n) \tag{3.11}$$

而第$(n+1)$个样本点只有x_{n+1}的值，问题是用下列主成分模型预测未观测到的值y_{n+1}，

$$x_i=\alpha_1+Af_i+\varepsilon_i \tag{3.12}$$

$$y_i=\alpha_2+b'f_i+\eta_i \tag{3.13}$$

$$i=1,\cdots,n+1$$

其中$\mathrm{cov}(\varepsilon_i,\eta_i)=0$，$\mathrm{cov}(\varepsilon_i)=\sigma^2 I$，$V(\eta_i)=\sigma_0^2$，而其余假设和模型（3.1）相同。上述问题在一系列论文（参见 Rao（1975，1976，1978，1987））以及 Rao 和 Boudreau（1985）中进行了研究。最近，模型（3.12－3.13）用于发展偏最小二乘法 （partial LS）（参见 Helland（1988）及其参考书目）。

这个问题有几个可能的处理方式。

1）假设 $\hat{f}_1,\cdots,\hat{f}_{n+1}$ 是仅用观测方程（3.12）得出的 $f_1,\cdots,f_{n+1}$ 的估计值。然后用（3.13）的前 n 个观测方程并假设 $\hat{f}_1,\cdots,\hat{f}_{n+1}$ 已知，用普通最小二乘法求 α_2 和 b 的估计值 $\hat{\alpha}_2$ 和 $\hat{b}$。最后用公式预测 y_{n+1}：

$$\hat{y}_{n+1}=\hat{\alpha}_2+\hat{b}'\hat{f}_{n+1} \tag{3.14}$$

2）令 $\hat{\alpha}_1$，$\hat{\alpha}_2$，$\hat{A}$ 和 $\hat{b}$ 是用（3.12）和（3.13）的前 n 个观测方程得出的 α_1，α_2，A 和 b 的估计值，然后假设 $\hat{\alpha}_1$ 和 $\hat{A}$ 已知，通过最小二乘法由方程

$$x_{n+1}=\hat{\alpha}_1+\hat{A}f_{n+1}+\varepsilon_{n+1} \tag{3.15}$$

估计 f_{n+1}。如果 $\hat{f}_{n+1}$ 是 f_{n+1} 的估计值，那么 y_{n+1} 就可以用

$$\hat{y}_{n+1}=\hat{\alpha}_2+\hat{b}'\hat{f}_{n+1} \tag{3.16}$$

预测。

3）用一个值（比如是 y）代替 y_{n+1}，使得方程（3.12－3.13）变完整。然后得出分块矩阵的奇异值分解

$$\begin{pmatrix} x_1:\cdots:x_n & x_{n+1} \\ y_1:\cdots:y_n & y \end{pmatrix}\left(I-(n+1)^{-1}ee'\right)=\ell_1 c_1 q_1'+\cdots+\ell_{p+1}c_{p+1}q_{p+1}'$$

其中 ℓ_i 取决于 y，并计算

$$S_r(y)=\ell_{r+1}^2(y)+\cdots+\ell_{p+1}^2(y) \tag{3.17}$$

最后用使（3.17）最小化的 y 值作为 y_{n+1} 的预测值。这个解可以通过图解或者用 Rao 和 Boudreau（1985）阐述的迭代算法得出。

4）另一种方法是将 f_i 看成是一个均值为零、协方差矩阵为 Γ 的随机变量，于是

$$\operatorname{cov}=\begin{pmatrix} x_i \\ y_i \end{pmatrix}=\begin{pmatrix} A\Gamma A'+\sigma^2 I & A\Gamma b \\ b'\Gamma A' & b'\Gamma b+\sigma_0^2 \end{pmatrix} \tag{3.18}$$

用（3.12）和（3.13）的前 n 个观测方程，得出 A，Γ，b，σ^2 和 σ_0^2 的估计值。Bentler（1983），Sörbom（1974）和 Rao（1983，1985）描述的方法可用于这一用途。那么 y_{n+1} 可以用

$$\hat{y}_{n+1}=\bar{y}+b'\Gamma A'\left(A\Gamma A'+\sigma_0^2 I\right)^{-1}\left(x_{n+1}-\bar{x}\right) \tag{3.19}$$

进行预测，其中 $\bar{y}=n^{-1}\Sigma y_i$， $\bar{x}=(n+1)^{-1}\Sigma x_i$ 而 b ， Γ， A 和 σ_0^2 用它们的估计值代替。

4.因子分析

4.1.一般讨论

在 FA 中， p 维的变量向量 x 被赋予一个随机结构

$$x=\alpha+Af+\varepsilon \tag{4.1}$$

其中 α 是 p 维向量，A 是 $p\times r$ 参数矩阵， f 是 r 维潜变量向量，称为共同因子，而 ε 是 p 维的变量向量，称为特殊因子(specific factors)，满足下列假定：

$$\mathrm{E}(\varepsilon)=0,\mathrm{cov}(\varepsilon)=\Delta \text{（是对角矩阵）}$$

$$\mathrm{E}(f)=0,\mathrm{cov}(f,\varepsilon)=0,\mathrm{cov}(f)=I \tag{4.2}$$

从（4.2）可以得出

$$\Sigma=\mathrm{cov}(x)=AA'+\Delta \tag{4.3}$$

要注意的是，当 $\Delta=\sigma^2 I$ ，（4.3）将变成（3.1）中讨论的主成分模型。以 x 的 n 个独立观测值 $x_1,\cdots,x_n$ 为基础，FA 中讨论的问题一般有:

1）表达式（4.3）拥有的最小 r 值是多少？
2）如何估计称为因子载荷矩阵的 A？
3）如何解释因子？
4）对于给定的个体,如何估计已知可观测 x 下的 f ？

应该注意，即使给定 r ，也并不保证方程（4.3）存在唯一的 A，方程（4.1）中的 f 也是如此。然而，目的是得出任一特解并考虑 A 和 f 变换的解释。有关 A 与 f 不可识别性以及因子旋转的讨论参见 Basilevsky（1994，pp.3５5-360，402-404），Jackson（1991，pp.393-396），Jolliffe（1986，pp.117-118）。

记 $X=(x_1,\ldots.,x_n)$，并计算

$$\bar{x}=n^{-1}Xe$$

$$S=(n-1)^{-1}X\left(I-n^{-1}ee'\right)X'$$

作为 α 和 Σ 的估计值。然后从 S 为出发点估计 A 和 Δ。最经常使用的是变量向量 x 多元正态性假设下的最大似然法（ML)。有许多计算机软件包可以估计因子个数 r 、因子载荷矩阵

A、特殊因子方差阵Δ（例如 SPSS，SAS，OSIRIS，BMD，COFAMM 等，它们也提供了 ML 估计以外的其他估计方法，同样可以计算用于解释的因子载荷旋转）。把A和Δ的 ML 估计值记为$\hat{A}$和$\hat{\Delta}$。

检验有r个共同因子假设的似然比检验准则是

$$-(n-1)\log\frac{|S|}{\left|\hat{A}\hat{A}'+\hat{\Delta}\right|} \tag{4.5}$$

在大样本中，它渐近服从自由度为$\left[(p-r)^2-p-r\right]/2$的$\chi^2$分布。这种检验在多元正态性假设下是有效的。将（4.5）中的乘数$(n-1)$用

$$n-1-\frac{2p+5}{6}-\frac{2r}{3} \tag{4.6}$$

代替，可以得到χ^2近似值的一个小改进。

另一种估计A和Δ的方法称为典型因子分析（CFA），是由 Rao（1955）发展起来的，它不需要任何分布假设。结果证明这种方法的解和 ML 估计相同。但是，（4.5）的χ^2检验要求多元正态性假设。

一般建议用计算机软件包提供的方法进行基于观测数据$x_1,\cdots,x_n$的多元正态性检验。有关正态性检验的部分参考书目有 Basilevsky（1994，4.6.2 节）和 Gnanadesikan（1977，5.4.2 节）。也可以对变量进行变换以达到正态性。但在这种情况下，因子结构是对变换后的变量而言的。

要注意的是，和 PCA 不同，如果使用无尺度的提取方法如 ML 和 FCA，FA 在尺度变量下，具有不变性，在这些情况下，可以从协方差矩阵或者相关矩阵的使用开始。如果用的是协方差矩阵而且量纲变动非常大，尺度因子将使得结果的解释变得复杂。此时，使用相关矩阵具有某些优点。当目标是比较组间因子结构时，使用协方差矩阵较适宜。

4.2.估计因子得分

在Σ的表达式中，使用A和Δ的估计值$\hat{A}$和$\hat{\Delta}$，可以估计测量指标x_i的第i个个体的因子得分f_i，由

$$\hat{f}_i=\hat{A}'\left(\hat{A}\hat{A}'+\hat{\Delta}\right)^{-1}(x_i-\bar{x}), i=1,\cdots,n. \tag{4.7}$$

（4.7）只不过是用估计值代替未知参数后f_i对x_i的回归。还有其他用于估计因子得分的表达式（参见 Jackson（1991，p.409））。

4.3.预测问题

考虑一个 $p+1$ 个变量 (x,y)，因子结构为

$$
\begin{aligned}
x &= \alpha + Af + \varepsilon \\
y &= \beta + a'f + \eta
\end{aligned}
\tag{4.8}
$$

其中 β 是一个纯量，a 是 r 维向量而 η 满足 $\mathrm{E}(\eta)=0$，$\mathrm{cov}(\varepsilon,\eta)=0$，$V(\eta)=\delta_{p+1}^2$。假设有 n 个个体的观测值 $(x_1,y_1),\cdots,(x_n,y_n)$ 以及仅有第 $(n+1)$ 个个体的观测值 x_{n+1}。问题是在给定所有其他观测值的基础上预测 y_{n+1}。考虑 $(p+1)$ 维变量向量的因子结构

$$
\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} A \\ a' \end{pmatrix} f + \begin{pmatrix} \varepsilon \\ \eta \end{pmatrix}
\tag{4.9}
$$

并使用观测值 $(x_1,y_1),\cdots,(x_n,y_n)$ 估计所有未知参数。假设 $\hat{\alpha}$，$\hat{\beta}$，$\hat{A}$，$\hat{a}$，$\hat{\Delta}$ 和 $\hat{\delta}_{p+1}^2$ 是用 CFA 或 ML 法得出的相应参数估计值。那么 y_{n+1} 对 x_{n+1} 的回归估计就是

$$
\hat{y} = \hat{\beta} + \hat{\alpha}'\hat{A}'\left(\hat{A}\hat{A}' + \hat{\Delta}\right)^{-1}\left(x_{n+1} - \hat{\alpha}\right).
\tag{4.10}
$$

在这种情况中，对参数 a，A 和 Δ 的估计，没有利用 x_{n+1} 提供的信息。

4.4.PCA 和 FA 之间的区别是什么？

在 PCA 中，没有对 p 维随机向量 x 设定任何结构。假设 $\mathrm{E}(x)=0$ 且 $\mathrm{cov}(x)=\Sigma$。我们希望用少量线性组合 $y=L'x$ 代替 x，其中 L 是秩为 r 的 $p\times r$ 矩阵。从而给定 y 时，x 的预测值（即 x 对 y 的回归）是

$$
\hat{x} = \Sigma L(L'\Sigma L)^{-1} y
\tag{4.11}
$$

而残差 $x-\hat{x}$ 的协方差矩阵是

$$
\Sigma - \Sigma L(L'\Sigma L)^{-1} L'\Sigma.
\tag{4.12}
$$

选取 L 作为最小化（4.12）的一个适当标准。选择 Frobenius 标准得到的解是

$$
L = (c_1 : \cdots : c_r)
\tag{4.13}
$$

其中 $c_1,\cdots,c_r$ 是 Σ 的前 r 个特征向量，如第 3 节解释的，这里的 $L'x$ 表示前 r 个主成分。目标是用少量变量尽可能解释 x 的整个协方差矩阵。

在 FA 中，给 p 维变量向量 x 的相关矩阵 R，赋予一个型如 $AA'+\Delta$ 的表达式。由于 Δ 是自由参数的对角矩阵，矩阵 A 最终由最小化 AA' 和 R 的非对角线元素之差来确定。因此，因子载荷矩阵是为解释观测变量之间的相关关系而设计的。变量中没有被因子解释的方差，不论大小，都被描述成特殊方差。在 PCA 中，重点更在于解释共同因子和特殊因子产生的

全部方差。因而，PCA 和 FA*的目标是不同的，解也是不相同的。

要点 1. 对 R 赋予型如 $AA'+\Delta$ 的表达式，对因子个数 r 施加了一个自动上界。因而，在一个既定的情况下，人们被迫用远小于可能对数据有影响的因子个数解释数据。在 Rao（1955）发展起来的 CFA 中，对共同因子的个数没有任何限制，但是这种方法要求从数据中提取主导因子的必要个数。开始时没有要求固定的因子个数，于是被视为一个估计问题而不是因子个数的假设检验问题。

要点 2. 有趣的是，在 FA 模型的公式中，仅使用了共同因子和特殊因子的二阶性质。然而，如果要求所有这些变量的分布是独立的，如下面 Rao（1969，1973）证明的定理所示，问题就变得更复杂了。

定理.令 x 是一个 p 维随机向量，具有线性结构 $x=Ay$，其中 y 是 q 维独立随机向量。那么 x 可以分解

$$x=x_1+x_2$$

其中 x_1 和 x_2 相互独立，x_1 具有实质上的唯一结构（$x_1=A_1y_1$，A_1 是唯一的且消除量纲，y_1 是数目固定的独立非正态向量），x_2 服从 p 元正态分布，具有不唯一的线性结构（$x_2=B_2y_2$，B_2 不要求唯一，而 y_2 是独立的一元正态变量组成的向量）。

根据这个定理，如果一些因子服从非正态分布，A_1 的唯一性自动指定因子变量个数的下限，这一下限可能与 p 无关。FA 模型仅考虑变量二阶性质的局限性需要进一步的研究。

4.5.套利定价理论模型（APT）

经典的 FA 模型由 Ross（1976）扩展为 APT 统计模型，类似于 Rao（1958，第 3 节，方程 9）的增长曲线模型。考虑一般的 FA 模型，用金融文献中的符号

$$R=\mu+Bf+u \tag{4.18}$$

其中 R 表示 N 种资产收益的 N 维向量，$\mu=\mathrm{E}(R)$，$\mathrm{E}(f)=0$，$\mathrm{E}(u)=0$，$\mathrm{E}(fu')=0$，$\mathrm{cov}(f)=\Phi$ 且 $\mathrm{cov}(u)=\Delta$，Δ 是一个对角矩阵。$N\times k$ 矩阵 B 是因子载荷矩阵。[在前面的各节中，p 用来表示 N 而 r 用来表示 k]。根据所作的假设

$$\Sigma=\mathrm{cov}(R)=B\Phi B'+\Delta \tag{4.19}$$

现在，对 μ 建立模型

$$\mu=R_f e+B\lambda \tag{4.20}$$

* 原书为 CA，有误。——译者注。

其中 R_f 是无风险资产的无风险收益。T 个时期的样本是

$$\left(R_1, R_{f1}\right), \cdots, \left(R_T, R_{fT}\right) \tag{4.21}$$

在（4.21）中，R_f 是已知的且随时间变化，而 λ 是 k 维未知参数向量，称为因子溢价。记 $r_t = R_T - R_{fT}e$，可以把第 t 个观测值的模型写成

$$r_t = B\left(f_t + \lambda\right) + u_t, t = 1, \cdots, T \tag{4.22}$$

这正是 Rao（1958）考虑的模型。r_t 的边际模型是

$$r_t = B\lambda + v_t, t = 1, \cdots, T \tag{4.23}$$

其中 $\operatorname{cov}\left(v_t\right) = \Sigma$。如果 B 和 Σ 已知，λ 的最小二乘估计量就是

$$\hat{\lambda} = \left(B'\Sigma^{-1}B\right)^{-1} B'\Sigma^{-1}\bar{r} \tag{4.24}$$

其中 $\bar{r} = T^{-1}\left(r_1 + \cdots + r_T\right)$。如果 B 和 Σ 未知，Roll 和 Ross（1980）以及 Rao（1958）提出，它们可以用 ML 或者适当的非参数方法进行估计，将具有无约束 μ 的模型（4.18）视为本文 4.2 节讨论的模型，并代入（4.23）。如果假设模型（4.18）中的 f 和 u 是多元正态分布，就可以基于观测值 $r_1, \cdots, r_T$ 写出所有未知参数 B, λ, Φ 和 Δ 的似然函数，并得出所有未知参数的 ML 估计值。还可以对 Σ 的设定进行似然比检验，即检验因子个数和 μ 的结构（4.20）。Christensen（1995）全部完成了这样的程序，其中用的是纽约证券交易所的数据。

5.结论

PCA 和 FA 都可以视为探测性数据分析的多元技术。两种分析的目标都是通过减少变量的个数以了解数据的结构，在某种意义上可以取代原始数据，而且通过图示和多元推断技术更容易进行研究。谨慎是必要的，因为需要进行许多决策，如减少变量的个数，判断简化变量集代表原始变量全体的充分性标准。

一些实践者将 PCA 和 FA 视为用于回答相同问题的可替代方法。有人认为，这两种方法已经发展成有用的数据分析工具，而且已经成为其他统计模型如聚类分析、判别分析、最小二乘回归、图形数据显示等等不可估量的帮助。如本文所讨论的，PCA 和 FA 中数据简化的目的是不同的。在 PCA 中，简化数据的目的是要依照整个协方差矩阵最大限度地近似原始数据的离散程度，而在 FA 中，重点在于解释原始变量之间的相关或者联系。两者的目标是不同的，因而必须就特定情况和数据分析目的来判断 PCA 和 FA 的适当性。虽然 PCA 和 FA 在探测性数据分析中的角色是清晰的，但精确使用估计的因子和主成分进行数据推断分析或者计划进一步研究的展示(laid out)，看起来不令人满意。

Schneeweiss 和 Mathes（1955）给出了因子得分和主成分彼此接近的条件。进行这种理论研究以及在个别的数据集中考察主成分与因子得分之间的实际差异，都是令人非常感兴趣的。

参考文献

Bartholomew, D. J. (1987). Latent Variable Models and Fact r Analysis. Oxford University Press, New York.

Basilevsky, A. (1994). Statistical Factor Analysis and Related Methods. Wiley, New York.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika 48, 493-517.

Benzecri, J. P. (1973). L'analyze des Donnes, Tome II, L'Analyse des Correspondences. Dunod, Paris.

Cattel, R. B. (1978). The Scientific Use of Factor Analysis in Behavioral and Life Science. Plenum Press.

Christensen, B. J. (1995). The likelihood ratio test of the APT with unobservable factors against the unrestricted factor model. Tech. Rept.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Ann. Eugen, London 7, 179-188.

Gnanadesikan, R. (1977). Methods for Statistical Analysis of Multivariate Observations. Wiley, New York.

Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis. Academic, London.

Helland, I. S. (1988). On the structure of partial least squares regression. Commun. Statist. Simula. 17, 581--607.

Hotelling, H. (1933). Analysis of a complex of statistical variable into principal components. Psychometrika 1, 27-35.

Jackson, J. E. (1991). A User's Guide to Principal Components. Wiley, New York.

Jolicoeur, P. and J. E. Mosiman (1960). Size and shape variation in the painted turtle, a principal component analysis. Growth 24, 339-354.

Joliffe, I. T. (1986). Principal Component Analysis. Springer-Verlag, New York.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. Proc. Roy. Soc. Edinburgh (A), 60, 64-82.

Pearson, K. (1901). On lines and planes of closest fit to a system of points in space. Philosophical Magazine 2, 6-th Series, 557-572.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. Psychometrika 20, 93-111.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. Biometrics 14, 1-17.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. Sankhya A 26, 329-358.

Rao, C. R. (1969). A decomposition theorem for vector variables with a linear structure. Ann. Math. Statist. 40, 1845-1849.

Rao, C. R. (1973). Linear Statistical Inference and its Applications, 2nd ed., Wiley, New York.

Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. Biometrics 31, 545-554.

Rao, C. R. (1976). Prediction of future observations with special reference to linear models. In: P. R.

Krishnalah, ed., Multivariate Analysis VI, North Holland, 193-208.
Rao, C. R. (1983). Likelihood ratio tests for relationships between covariance matrices. In: S. Karlin, T. Ameniya and L. A. Goodman, eds., Studies in Economics, Time Series and Multivariate Statistics. Academic, New York, 529--543.
Rao, C. R. and R. Boudreau, (1985). Prediction of future observations in factor analytic type growth model. In: P. R. Krishnaiah, ed., Multivariate Analysis VI. Elsevier, Amsterdam, 449-466.
Rao, C. R. (1987). Prediction of future observations in growth curve models. J. Statist. Science 2, 434-471.
Rao, C. R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. Q esfii6 19, 23-63.
Roll, R. and S. A. Ross (1980). An empirical investigation of the arbitrage pricing theory. J. Finance 35, 1073-1103.
Ross, S. A. (1976). The arbitrage theory of capital asset pricing. J. Econom. Theory 13, 341-360.
Schneeweiss, H. and Mathes, H. (1995). Factor analysis and principal components. J. Multivariate Analysis 55, 105-124.
S bom, D. (1974). A general method for studying differences in factor means and factor structure between groups. British J. Math. Statist. Psych. 27, 229-239.
Spearman, C. (1904). General intelligence, objectively determined and measured. Am. J. Psych. 15, 201-293.
Stone, R. (1947). An interdependence of blocks of transactions. J. Roy. Statist. Soc. (Supple), 8, 1-32.
Wegman, E. J, D. B. Carr and Q. Luo (1993). Visualizing multivariate data. In: C. R. Rao, ed., Multivatiate Analysis: Future Directions. North Holland, 423--466.